

WHAT IS CLAIMED IS:

1. A service distribution device for distributing services among a plurality of servers on a network to balance the server loads, comprising:

5 a packet capture device capturing packets transmitted through the network;

a server identifier recording information pertaining to the captured packets into a server log for each server;

a service identifier recording information pertaining to the captured packets into a service log for each service;

10 a server modeling module setting up the simulation model for each server from the server records;

a service modeling module setting up the simulation model for each service from the service records;

15 a simulator reading in the server model and the service model and running each simulation; and

a server selection module selecting and specifying an optimum server to distribute services to based on a simulator result.

2. The service distribution device of claim 1, further comprising a packet relay

20 device obtaining packets using a packet capture module mounted on said packet relay device, which relays packets between a client and the servers.

5

5

5

0

5

0

0

0

5. The service distribution device of claim 1, wherein said simulator performs a simulation using the server model and the service model and generates a mean value or a median value of a session time for the specific service.

5

6. The service distribution device of claim 1, wherein said server selection module determines a standard value using an output of a single simulation run for each service by said simulator, and determines that a high load state exists if a difference between, or the ratio of, the standard value and the output of the simulation of a plurality of sessions exceeds a pre-determined threshold.

7. The service distribution device of claim 6, wherein when said server selection module receives a server distribution query, said server selection module sets a server permission to be a starting frequency of the session that will cause a high load state for the service in question for each server, and specifies a server having the biggest difference between the session starting frequency and the permission as a server for distribution.

8. The service distribution device of claim 6, wherein when said server selection module receives a distribution server query, said server selection module runs a simulation for a service in question for each server and specifies a server for which a result of a ratio for which β multiplied by the standard value is less than or equal to γ .

9. The service distribution device of claim 6, wherein when said server selection module receives a distribution server query, said server selection module runs a simulation for a service in question for each server and specifies as a distribution server, a server for which a result of ratio for which β multiplied by the standard value is smallest.

5

10. The service distribution device of claim 4, wherein said service modeling module categorizes each session transmission as a connection request and response, and a command transmission, a data transmission, a response, and an end, and calculates the parameters for each session transmission based upon category.

11. The service distribution device of claim 7, wherein the permissions of each of the servers are taken as weighted values of a service distribution or relative ratios of the permissions are taken as server distribution ratios.

12. A service distribution device for distributing services among a plurality of servers to balance server loads, comprising:

a server modeling module generating a simulation model for each server and a service modeling module generating a simulation model for each service;

a simulator reading the server models and the service models and running a plurality of simulations; and

a server selection module determining which servers have low loads based on results of the simulations and selecting the servers with low loads to receive the services.

loads, comprising:

generating a simulation model for each server and each service;

running a plurality of simulations using the server and service models; and

determining which servers have low loads based on results of the simulations and selecting the servers with low loads to receive the services.

process of:

generating a simulation model for each server and each service;

running a plurality of simulations using the server and service models; and

and selecting the servers with low loads to receive the services.